DOCUMENT RESUME

ED 038 673 CG 005 218

AUTHOR Archambault, Francis X., Jr.
TITLE A Computerized Approach to Scoring Verbal Responses to the Torrance Tests of Creative Thinking.
INSTITUTION Boston Univ., Mass.
PUB DATE [70]
NOTE 11p.

EDRS PRICE EDRS Price MF-$0.25 HC-$0.65
DESCRIPTORS *Computer Oriented Programs, Computers, *Creative Ability, *Creative Thinking, Creativity Research, Elementary School Students, Students, Testing, *Test Results, Tests, *Test Scoring Machines

ABSTRACT

This paper describes a study of a computerized approach to scoring the Torrance Tests of Creative Thinking (TTCT). A total of 153 students from grades four through seven were involved, 100 in a developmental sample on which the computorized scoring procedures were developed, and a cross validation sample composed of the remaining 53. This research was limited to three of the seven subtests of the TTCT. Subjects' responses to each of the activities are scored for fluency, flexibility, and originality. The fluency score is defined as the total number of relevant responses given; flexibility as the number of different clusters of responses, originality was scored based on three dictionaries, with originality weights of zero, one, and two. The step-wise multiple regression technique was employed to maximize the prediction of each subject's score for each activity of the TTCT. The prediction of fluency was the most accurate. However, with some corrections, both flexibility and originality results were improved. It appears that creativity, as defined by Torrance can be judged accurately by a computer. (KJ)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

A COMPUTERIZED APPROACH TO SCORING VERBAL RESPONSES TO THE TORRANCE TESTS OF CREATIVE THINKING

Francis X. Archambault, Jr.
Boston University

ED038673

Since the last decade when Guilford (1950) called attention to the virtual neglect of the concept of creativity by American researchers, there has been an enormous expansion of interest and research in the nature of this higher mental process. A myriad of problems and controversies have surrounded work in the area of creativity, but one of the most pressing issues continually has been the search for valid and reliable means of measuring creative performance.

The recent publication of the Torrance Tests of Creative Thinking (Torrance, 1966) in many respects may be regarded as a breakthrough in the area of creativity measurement. Based on nearly nine years of research and development by Torrance and his colleagues, the tests represent a pioneering venture in that they provide the researcher and educational practitioner with a functional instrument for measuring creative potential in children, adolescents, and adults. In spite of the relatively high level of development of the Torrance instruments, certain technical problems related to levels of training on the part of the scorers may act as a deterrent to their widespread use. At least one reviewer (Hoepfner, 1967) has called attention to these problems and also has suggested that the time required to score the test battery may be a relatively long affair. These

CG005218

shortcomings may be dismissed, however, by using a computer to score the verbal responses to the TTCT for, unlike humans, the computer functions as a perfectly reliable judge which does not suffer from fatigue or lapses of attention. Moreover, the computer might perform this service with savings of both time and money.

To determine the effectiveness of such a computerized approach a sample of 153 pupils from grades 4, 5, 6, and 7 in six Central New York State public school systems was employed. These 153 subjects were randomly assigned to a developmental sample of size 100, on which the computerized scoring procedures were developed, and a cross-validation sample composed of the remaining 53 subjects. (Mosier, 1951).

Each of these subjects was administered the TTCT, Verbal Form A, but the present research dealt solely with the open-ended responses to three of the seven activities or subtests included in the battery. The activities considered were the Ask and Guess subtests (Activities 1, 2, and 3) in which subjects ask questions about a drawing and make guesses about the causes and consequences of a pictured event.

The subjects' responses to each of these activities are scored by human judges for Fluency, Flexibility, and Originality. Fluency is, according to Torrance, the total number of relevant responses given for each activity; Flexibility is the number of different categories of responses or the number of shifts in response emphasis for each of the subtests; and Originality is a

measure of the infrequence of each response. The Originality score for each activity is the sum of the Originality scores for each of the individual responses.

Using the scoring procedure set forth in the Directions Manual and Scoring Guide of the TTCT four trained human judges (Archambault, 1969) scored the responses of the 153 subjects. The separate judge scores were then pooled to obtain criterion measures against which the performance of the computerized approach could be guaged. The pooled reliabilities of these judges (Winer, 1962, pp. 124-132) are shown in Table I. As evident from the table, the reliabilities are all extremely high, with the possible exception of Activity 3, Originality.

To perform the computerized scoring of the data it was first necessary to transcribe the responses of each subject into machine readable form. This was accomplished by keypunching the responses on standard IBM cards, one response to a card. Since no corrections in spelling, punctuation, grammar, etc., were made on the original copy, the keypunched data were an exact duplicate of the responses given in the test booklets. The actual scoring of the test was performed by Fisher's (1968) SCORTXT program, a system consisting of a main program and nine subroutines currently operating under the IBM 360 OS system. In using the Fisher program two separate scoring strategies were employed, sometines in concert. The first strategy was modeled directly after the manual scoring procedure developed by Torrance. The second involves the use of various actuarial measures which have proven valuable in related

Table 1

RELIABILITY ESTIMATES FOR FOUR JUDGES FOR FLUENCY, FLEXIBILITY, AND ORIGINALITY OF ACTIVITIES 1, 2, and 3 OF THE TORRANCE TESTS OF CREATIVE THINKING, VERBAL FORM A USING ANALYSIS OF VARIANCE

	Total Sample	Developmental Sample	Cross-Validation Sample
Activity 1, Fluency	.99	.99	.99
Activity 1, Flexibility	.98	.98	.98
Activity 1, Originality	.81	.81	.79
Activity 2, Fluency	.95	.96	.95
Activity 2, Flexibility	.93	.93	.93
Activity 2, Originality	.80	.84	.71
Activity 3, Fluency	.93	.94	.91
Activity 3, Flexibility	.92	.93	.90
Activity 3, Originality	.66	.73	.52

research by a number of investigators (Page and Paulus, 1968; Marcotte, 1969; McManus, 1968). Since the responses were judged at separate times for Fluency, Flexibility, and Originality and since the scoring strategy used is dependent on whether Fluency, Flexibility, or Originality is being assesed the method used for each of these will be described separately.

As mentioned previously, the Fluency score for each activity is defined as the total number of relevant responses given. It was hypothesized that the Fluency score could be determined without assessing the relevance of the individual responses, and that, because of this, simple actuarial measures could be used to predict Fluency. Following this hypothesis, students' responses were reduced by SCORTXT to a series of counts or frequency scores on a variety of variables, a listing of which is given in Figure I. These variables were then used in a step-wise multiple regression analysis to predict the Fluency score.

The Flexibility score for each activity is defined as the number of different clusters of responses or the number of shifts in response emphasis. For each activity of the TTCT, Torrance has isolated categories into which the responses might fall. Twenty-two such categories have been isolated for Activity I, while for Activities 2 and 3, 21 Flexibility clusters have been determined. For each of these categories a dictionary of entries to be used in the computerized scoring procedure was built. The dictionaries were constructed by analyzing the model responses given by Torrance for key words and phrases and then isolating synonyms of these key words and phrases

Number of Question Marks

Number of Commas

Number of Periods

Number of Words of Length One

Number of Words of Length Two

Number of Words of Length Three

Number of Words of Length Four

Number of Words of Length Five

Number of Words of Length Six

Number of Words of Length Seven

Number of Words of Length Eight

Number of Words of Length Nine

Number of Words of Length Ten

Number of Words

Number of Sentences

Number of Paragraphs

Average Word Length

Average Sentence Length

Average Paragraph Length

Standard Deviation of Word Length

Standard Deviation of Sentence Length

Standard Deviation of Paragraph Length

Third Moment of Word Length

Fourth Moment of Word Length

FIGURE I

ACTUARIAL VARIABLES INCLUDED IN PREDICTION EQUATIONS
FOR FLUENCY, FLEXIBILITY, AND ORIGINALITY

in Roget's International Thesaurus (1962) and Soule's Dictionary of English Synonyms (1966). The responses ofthe students were then analyzed by SCORTXT performing a word/phrase lookup to determine how many categories were used. In addition, since high correlations were found between some of the actuarial measures and the Flexibility criteria, the variables listed in Figure I were again used in the analysis. These data, both the category counts and the actuarial scores, were used in the multiple regression analysis to predict the Flexibility scores.

For scoring Originality three dictionaries were constructed, based on the possible Originality weights which the response might receive. The first dictionary consisted of all zero weight entries listed in the scoring manual developed by Torrance along with the synonyms of these entries extracted from the Flexibility dictionaries already constructed. A similar procedure was followed for the construction of the second dictionary comprised of entries for which the Originality weights were one. The remaining Flexibility entries were then included in the Originality dictionary whose entries had weights of two. This procedure was followed for each of the three Activities. As with Flexibility, scores on the actuarial variables were used in the development of prediction equations.

As indicated previously, the step-wise multiple regression technique was employed to maximize the prediction of each subject's scores for each Activity of the TTCT. Since nine scores were predicted for each individual, that is, a Fluency, Flexibility, and Originality score for each of the three Activities,

nine separate analyses were performed yielding nine different prediction equations. The results of these analyses are summarized on Table 2.

Of the three scores that would be predicted for each activity it was hypothesized that the prediction of Fluency would be the most accurate. The results summarized in Table 2 support this hypothesis. That the multiple-R's would be so high had not been expected, however, since no scheme for the determination of the appropriateness of the responses was included in the scoring procedure. Similarly, the size of Mult-R's obtained in the prediction of both Flexibility and Originality were much higher than had been anticipated.

For the prediction of Flexibility and Originality, it was hypothesized that the variable "category counts" would be the most important predictor, since the counts were derived in accordance with Torrance's scoring norms. However, this was true only for the prediction of the Activity 1, Originality scores. For the prediction of the Flexibility scores of Activities 1 and 2 and the Originality score of Activity 2, "category counts" was the sixth best predictor; for Activity 3, Flexibility, it was the twelfth best predictor; and for Activity 3, Originality, the variable was not entered until the 24th step of the regression analysis. A number of explanations might be given for these results, but the explanation advanced earlier by Dieter Paulus (i.e., that Fluency is a necessary condition for Flexibility and Originality) appears the most appropriate.

In cross validation, the multiple-R's for Fluency held up very well, but sizeable shrinkage was found for the multiple-R's of the Flexibility and

Table 2

SUMMARY OF RESULTS OF STEP-WISE
MULTIPLE REGRESSION ANALYSIS
BOTH DEVELOPMENTAL AND CROSS-VALIDATED

Criterion	Multiple-R Developmental (N=100)	Multiple-R Cross-Validated (N=53)	Cross-Validated Multiple-R Correlated for Attenuation (N=53)
Activity 1, Fluency	.97**	.89**	.90**
Activity 1, Flexibility	.91**	.71**	.72**
Activity 1, Originality	.93**	.74**	.83**
Activity 2, Fluency	.93**	.88**	.90**
Activity 2, Flexibility	.87**	.68**	.71**
Activity 2, Originality	.83**	.75**	.89**
Activity 3, Fluency	.95**	.88**	.92**
Activity 3, Flexibility	.85**	.56**	.59**
Activity 3, Originality	.91**	.72**	.99**

** Significant at .01 level

Originality dimension. However, when adjustments were made for the lack of perfect reliability in the criteria (i.e., the so-called "correction for attenuation") significant increases in the correlations were found for both of these dimensions. Moreover, if fewer predictors were used in the development of the regression equations, as seems appropriate from the results obtained, the correlation found in cross validating the results would have been higher.

It appears, then, that creativity, as defined by Torrance, can be judged accurately by a computer. Further, it appears that the use of a computer to score open-ended responses to other standardized tests may be appropriate and should be investigated.

REFERENCES

Archambault, F. X. A Computerized Approach to Scoring Verbal Responses to the Torrance Tests of Creative Thinking. Unpublished Doctoral Dissertation, University of Connecticut, 1969.

Fisher, G. A. The SCORTXT program for the analysis of natural language. Unpublished manuscript: Bureau of Educational Research, University of Connecticut, 1968.

Guilford, J. P. Nature of human intelligence. New York: McGraw-Hill, 1967.

Hoepfner, R. Review of Torrance tests of tests of creativity. Journal of Educational Measurement, 1967, 4, 191-193.

McManus, J. F. Computer evaluation of college history examinations by actuarial strategies. Unpublished Doctoral dissertation, The University of Connecticut, 1968.

Marcotte, D. R. A computerized contingency analysis of content graded essays. Unpublished Doctoral dissertation, The University of Connecticut, 1969.

Mosier, C. I. The need and means of cross-validation. Educational and Psychological Measurement, 1951, 11, 5-28.

Page, E. B. & Paulus, D. H. The analysis of essays by computer. Final Report. Project No. 6-1318, Contract No. OEC-16-001318-1214. U. S. Department of Health, Education and Welfare. Storrs, Conn.: The University of Connecticut, 1968 (mimeo.).

Roget, P. (ed.). Roget's International Thesaurus. New York: Crowell, 1962.

Soule, R. Soule's dictionary of English synonyms. Boston: Little , Brown, & Co., 1966.

Torrance, E. P. Torrance Tests of creative thinking. Princeton, New Jersey: Personnel Press, 1966.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.